

# Automated Operative Step Identification in Peroral Endoscopic Myotomy

Thomas M. Ward, MD <sup>1</sup> Daniel A. Hashimoto, MD, MS <sup>1</sup> Yutong Ban, PhD <sup>1</sup> Guy Rosman, PhD <sup>2</sup> Daniela L. Rus, PhD <sup>2</sup> Ozanan R. Meireles, MD <sup>1</sup>

<sup>1</sup>Surgical AI and Innovation Laboratory, Massachusetts General Hospital <sup>2</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

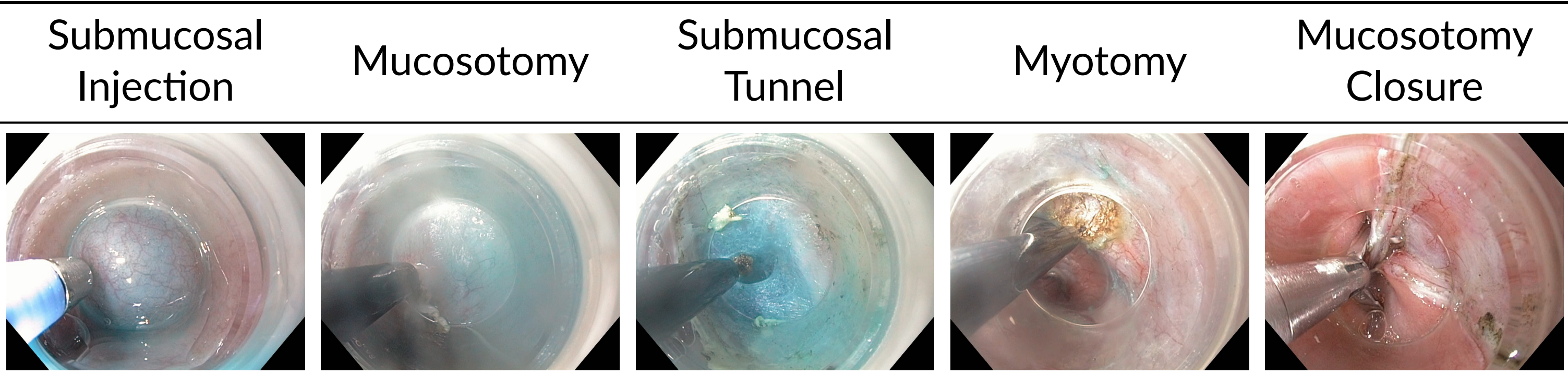
## Artificial intelligence and computer vision in surgery

We are in the midst of an artificial intelligence (AI) and computer vision (CV) revolution. In the medical field, CV algorithms can perform at levels similar to pathologists [3], radiologists [5], and dermatologists [1]. However, there have been relatively few applications of AI and CV in surgery. Current work focuses on automated operative step recognition, with work done in laparoscopic variants of cholecystectomy (86.7% accuracy [6]), sleeve gastrectomy (85.6% accuracy [2]), and sigmoid colectomy (91.9% accuracy [4]). We applied computer vision techniques for the first time on endoscopic procedures to identify surgical steps of intra-operative Peroral Endoscopic Myotomy (POEM) videos.

## How does a machine learn to identify surgical steps?

### Data and Annotations

50 POEM videos were collected from Massachusetts General Hospital (Boston, USA) and Showa University Hospital (Tokyo, Japan). AI algorithms train on video frames that surgeons have labeled with the correct operative step (“ground truth”). Surgeons labeled the POEM videos with the following “ground truth” steps:



### Deep Neural Networks

A deep learning computer vision model was then trained on the videos. The model’s architecture combined two neural networks:

1. Visual Model: Convolutional Neural Network (CNN) (Resnet18 in practice)
2. Temporal Model: Long Short-Term Memory (LSTM) Neural Network

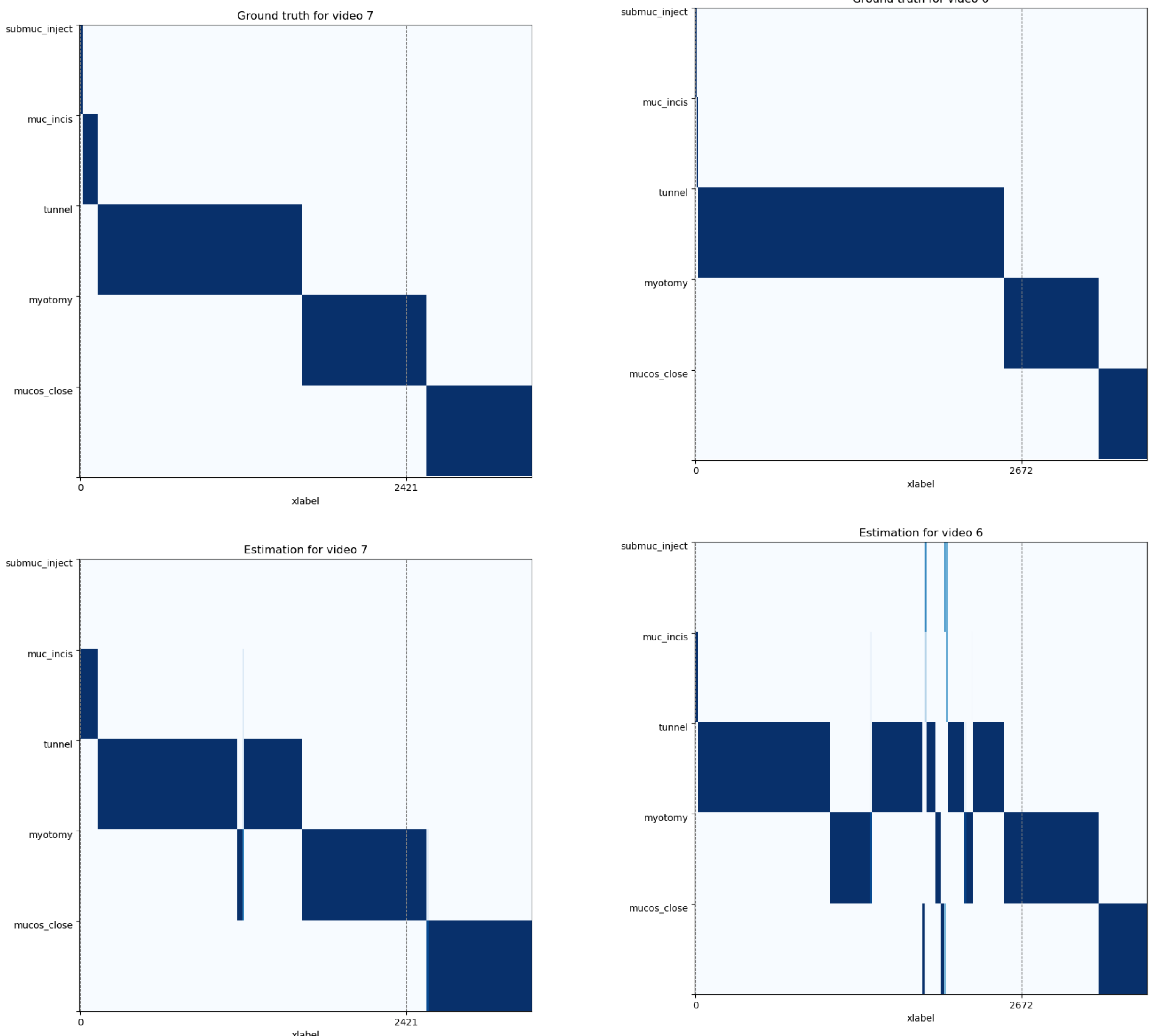
It trained on 30 annotated videos, learning both visual and temporal features to classify operative steps. The model’s inference capability was tested on the remaining 20 videos. Overall and per-step AI model inference performance was evaluated compared to surgeon-labeled steps with the following measures (calculated with the *caret* package in R 3.4.1):

1. Precision (correctly inferred frames/total number of frames inferred)
2. Recall (correctly inferred frames for a step/step’s total number of frames)
3. F-score (Harmonic mean of precision and recall)

## Step identification is 87.6% accurate

|            | Submucosal Injection | Mucosotomy | Submucosal Tunnel | Myotomy | Mucosotomy Closure | Overall      |
|------------|----------------------|------------|-------------------|---------|--------------------|--------------|
| Precision  | 0.667                | 0.837      | 0.955             | 0.791   | 0.848              | 0.885        |
| Recall     | 0.361                | 0.602      | 0.840             | 0.945   | 0.971              | <b>0.876</b> |
| F1         | 0.468                | 0.700      | 0.894             | 0.861   | 0.906              | 0.875        |
| Prevalence | 0.006                | 0.044      | 0.513             | 0.278   | 0.159              | 1            |

**Table 1.** The model did well in step identification across all steps other than “Submucosal injection,” which only occurred in 0.6% of frames. The overall metrics were calculated from a prevalence weighted-average of each step. Overall accuracy (frames correctly inferred/total frames) was 87.6% (95% CI 87.4% to 87.9%).



(a) Straightforward case with few frame mis-classifications.

(b) Case with tortuous esophagus causing tunneling difficulties (mucosal perforation requiring closure) that led to machine confusion.

**Figure 1.** “Surgical fingerprints” for two representative cases that show annotated ground truth (top) and the model’s inferred step (bottom).

## What confused the machine?

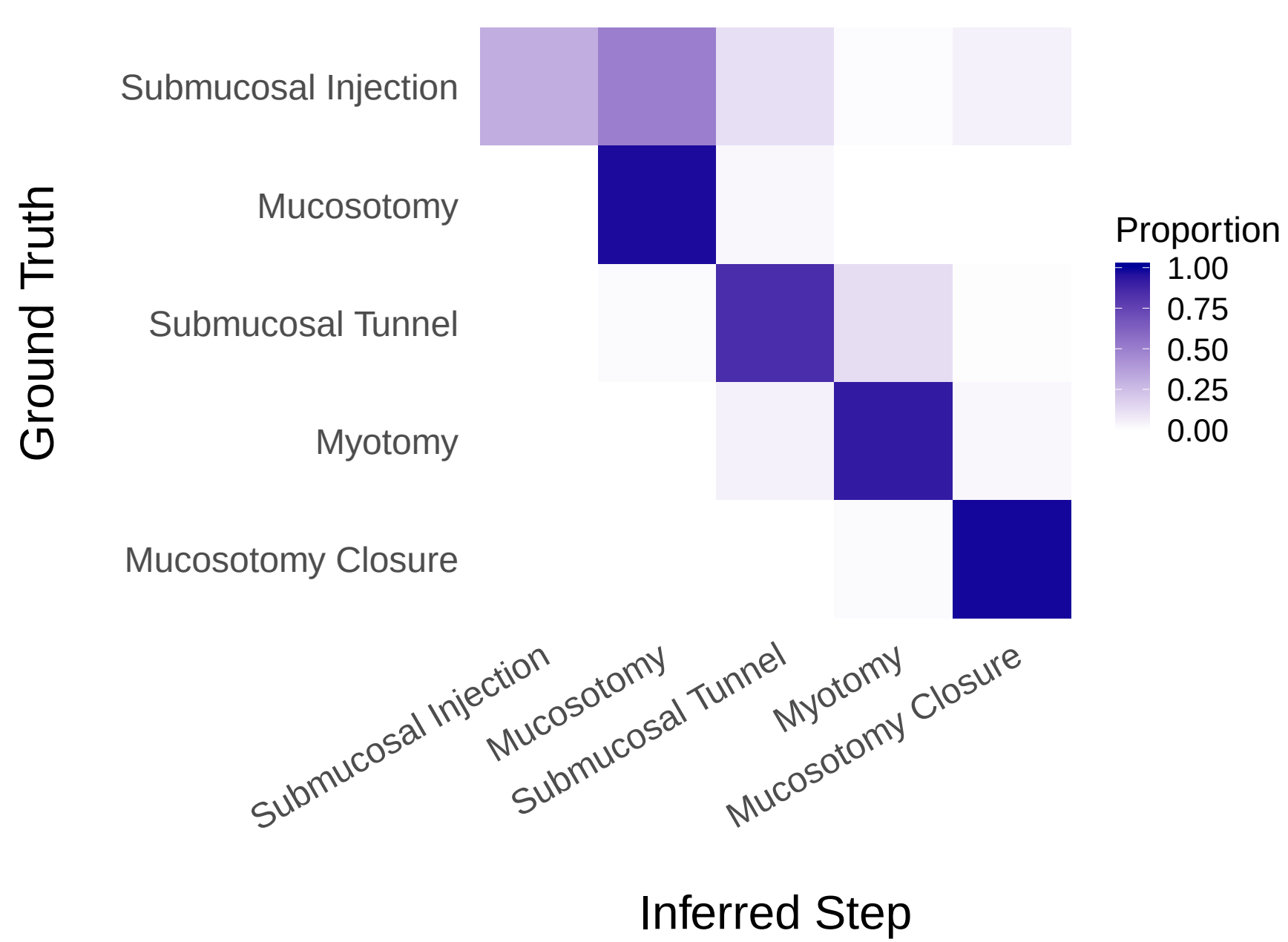


**Figure 2.** Confusion matrix that shows the proportion of inferred steps for each ground truth step. Submucosal injection was poorly inferred (classified as other steps more often than itself), likely due to few training examples for the machine to learn from given its low prevalence. “Submucosal tunnel” and “Myotomy”, visually similar tasks, had some shared mis-classification.

## Implications for Surgical AI

- Accurate automated operative step identification is possible for POEMs.
- AI and CV techniques perform well in endoscopic procedures in addition to their previous successes in laparoscopic procedures.

## Future directions

- Improve model’s classification through an increase in training data and more granular annotation schema to capture further surgical phenomena.
- Develop complication prediction and clinical decision support models.
- Correlate video findings to post-operative outcomes.

## References

- [1] Esteva et al. *Nature*. Feb. 2017.

[2] Hashimoto et al. *Annals of Surgery*. Sept. 2019.

[3] Hollon et al. *Nature Medicine*. Jan. 2020.
- [4] Kitaguchi et al. *Surgical Endoscopy*. Dec. 2019.

[5] Rajpurkar et al. Nov. 2017.

[6] Yengera et al. *arXiv*. May 2018.

## Acknowledgments

This research was supported by a 2017 research award from the Natural Orifice Surgery Consortium for Assessment and Research (NOSCAR). Many thanks to Dr. Haruhiro Inoue from Showa University Northern Yokohama Hospital for providing additional POEM videos and his pioneering of the POEM procedure.